工业控制网络通信异常检测的改进鱼群算法优化方法 *

陈万志 [1]，唐　雨 [1]，张　静 [2]

(1. 辽宁工程技术大学 电子与信息工程学院，辽宁 葫芦岛 125105; 2. 渤海装备辽河重工有限公司，辽宁 盘锦 124010)

摘　要：针对工业控制网络中典型的攻击类型，提出一种利用深度学习预测工控网络通信异常的方法。首先，利用主成分分析方法对原始数据降维，消除原始数据集的相关性；其次，构建人工神经网络并利用万有引力搜索算法中粒子惯性质量计算思想改进的鱼群算法来优化极限学习机的输入权值和阈值。测试实验结果表明，异常检测的准确率有所提升，同时有效地缩短了检测时间，实现了利用深度学习预测工控网络通信异常的行为。

关键词：工业控制网络；主成分分析；极限学习机；异常检测；人工鱼群算法；万有引力搜索算法

中图分类号：TP393.08　　doi: 10.3969/j.issn.1001-3695.2018.01.0099

Improved method of optimal fish swarm optimization for industrial control network communication anomaly detection

Chen Wanzhi[1], Tang Yu[1], Zhang Jing[2]

(1. *School of Electronic & Information Engineering Liaoning Technical University*, *Huludao Liaoning* 125105, *China*; 2. *China Petroleum Liaohe Equipment Company*, *Panjin Liaoning* 124010, *China*)

Abstract: Aiming at typical attack types of industrial control networks, this paper proposed a method of predicting communication anomalies in industrial networks using deep learning. First, the principal component analysis of the raw data reduction and eliminated the correlation between the original data set. Secondly, build artificial neural networks and to optimize the input weights and threshold limits the use of machine learning. The fish swarm algorithm was improved by the idea of particle inertia mass calculation in the gravitational search algorithm. The test experiment results show that the accuracy of anomaly detection is improved, and the detection times are effectively shortened. And realizes the purpose of making use of the depth learning to predict the abnormal behavior of communication in industrial networks.

Key words: industrial control network; principal component analysis; extreme learning machine; anomaly detection; artificial fish swarm algorithm; gravitation search algorithm

0　引言

工业控制系统对实时性要求高，没有充足的升级资源和安保功能，并且必须具有容错能力，不能停机或重启[1]。工控系统中的监控网络用来控制信息和操作信息的通信问题，通常采用传输层的 TCP、UDP 协议。然而传输层的协议未考虑安全性，极易受到拒绝服务攻击、中间人攻击、内部人攻击和端口扫描攻击等。传统的网络入侵检测在流量过滤和监测时存在细粒度过大、协议类型不兼容和在复杂的工业环境中无法防御中间人或内部人攻击[2]等问题，因此不能将传统入侵检测技术套用在工业环境的入侵检测中，需要制定针对工业控制系统环境和协议类型的入侵检测技术。

工业和信息安全领域的国内外学者现在正针对工业控制系统的异常检测进行研究，但是对于异常检测的相关研究尚处于起步阶段。尚文利等人[3]提出了一种基于 OCSVM 算法的入侵检测模型，并将粒子群算法应用其中，提高了系统的准确性。Yang 等人[4]提出将误用与异常结合的入侵检测机制，首先根据入侵行为特征库快速识别未授权访问等已知攻击；然后采用基于神经网络的异常检测机制，实现对未知攻击的入侵检测。这种方法可大幅度地提升 IDS 检测效率。侯重远等人[5]提出了基于概率主成分分析法的工业网络流量异常检测方法，建立了工业控制网络流量矩阵的概率主成分分析法模型，有效降低了误

收稿日期：2018-01-27；**修回日期：**2018-03-20　**基金项目：**辽宁省教育厅服务地方类项目（LJ2017FAL009）；辽宁工程技术大学博士启动基金资助项目（2015-1147）

作者简介：陈万志（1977-），男，辽宁阜新人，副教授，博士，主要研究方向为人工智能、计算机过程控制、物联网工程、信息安全等（chenwanzhi@lntu.edu.cn）；唐雨(1994-)，女，辽宁大连人，硕士研究生，主要研究方向为人工智能、信息安全；张静(1980-)，女，江苏徐州人，技术员，工程师，学士，主要研究方向为电气自动化、工业控制..

报率。但这种方法仅在攻击流量特征与正常业务类型差异较大的情况下才具有指导意义。Zhou 等人[6]通过对工控系统通信行为的特征数据的提取，组成机器学习的训练样本集合来测试样本集，利用机器学习方法建立异常行为入侵检测模型。 Aburo mman 等人[7]提出一种新的组合 SVM(support vector machine)-KNN(K-nearestneighbor)-PSO(particle swarm optimization)方法的入侵检测系统，利用 PSO 算法对权重进行优化，组合 SVM 和 KNN 这两种分类方法来获取更好的入侵检测精度。但该方法的设计与应用没有充分的考虑工控系统实时性的影响。

本文重点阐述了主成分分析算法(principal component anal ysis,PCA)、基于万有引力搜索算法(gravitation search algorithm, GSA)的人工鱼群算法(artificial fish swarm algorithm,AFSA)以及人工神经网络在工业控制网络通信异常检测方面的研究工作。通过 PCA 对原始数据进行预处理，消除变量之间的冗余性和相关性，较大程度上为人工神经网络的训练奠定了较好的数据基础，有利于提升训练结果的精度和稳定性。改进的人工鱼群算法避免了神经网络初始输入权值和阈值的随机性，同时有效地提高了算法的自适应能力和优化精度。通过对神经网络的初始输入权值和阈值进行优化后，使得 ELM 模型具有更高的预测精度，同时减少盲目寻找产生的训练时间的延长，使得工控网络通信异常行为预测模型有更好的泛化性能。因此 PCA-GSA-AFSA-ELM (extreme learning machine) 的工控网络通信异常行为预测模型可以对异常行为的预测进行快速高效的预测。

1　工业控制网络异常检测的原理

1.1　通信异常检测思想

近年来，工业控制系统为了便于企业管理，将上层管理系统与底层工业控制网络互连互通，用户通过企业管理信息网中的计算机 Windows 相关应用程序监控工业控制网中的工业设备运行参数，以及获取服务器上的生产数据。

因此，依据其特点采用如图 1 所示的体系结构构建异常检测模型，在管理信息网络中用户客户端与 Web 服务器间的路由器上加入主成分分析方法、改进的鱼群算法以及 ELM 神经网络的深度学习检测环节，利用通信数据特征，检测管理网络与服务器的通信异常行为，进一步提升准确率，有效缩短检测时间。

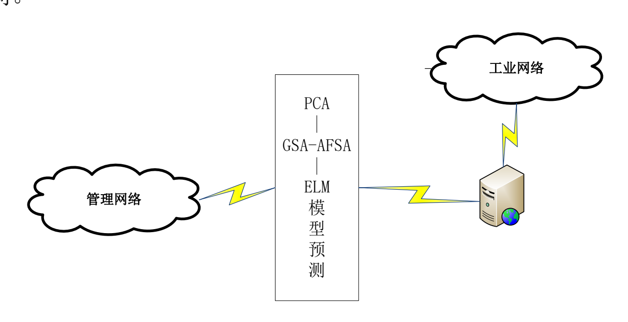


图 1　系统架构模型

1.2　通信异常行为的特征选择

通信数据中异常通信数据情况较为单一、样本数量较少。根据目前工业控制网络环境，本文针对以下攻击类型进行分析。

1)DoS 攻击　由网络协议本身的安全缺陷造成，最常见的 DoS 攻击是利用合理的服务请求来占用过多的服务资源，致使服务超载，无法响应其他的请求。这些资源包括网络带宽、文件系统空间容量、开放的进程、内向的连接等。这种攻击会导致系统资源的匮乏。设法远程或本地访问通信基础设施的攻击者可以通过淹没与虚假流量的关键链路或者通过使关键网络设备的计算资源饱和来启动拒绝服务（DoS）攻击，如路由器或计量现场设备。这种攻击导致来自现场设备的实时测量数据被延迟或最差丢弃[8]。

2)端口扫描　端口扫描攻击是远程用户(R2L)攻击的第一步。其目的是知道哪些系统端口是开放的。这是基于网络入侵的基本步骤。通过这次攻击，攻击者可以发现一些潜在的漏洞，并通过该漏洞侵入系统。IDS 可以使用源地址、目的地址和端口号识别数据包的模式。但是潜行扫描攻击（如慢速端口扫描）很难找出，因为缓慢的端口扫描攻击与正常活动非常相似[9]。

3)中间人攻击　通常分布在一些通信线路没有很好的物理保护的地方。攻击者可以通过选择性地删除或修改从现场设备（控制器）发送的传感器数据（控制信号）来发起中间人攻击，从而损害可用性和完整性的消息交换。重播攻击是中间攻击的另一种形式：嗅探通信通道的攻击者可以复制测量数据或控制命令，然后转发它们。中间人对测量数据的攻击主要是在攻击持续存在的情况下有效。这是因为系统是动态系统，即测量数据被新的一组测量持续刷新。因此，单个中间人攻击的效果可以忽略不计，特别是对于每秒刷新几次的同步测量；相反，对控制信号的单一攻击可能是灾难性的[10]。

4)内部攻击　内部攻击通常分为两种情况：a)内部人员使用他们的合法访问来执行与安全策略相违背的某些操作，如敏感数据泄露给某些第三方，若此时访问资源被允许和阻止时，则可能发现违规操作源；B)内部人员以访问控制的安全策略的方式，使用他们的访问来扩展其权限，当用户可能具有登录特定系统的合法能力，并滥用该权限以获取对系统的非法级别访问[11]。

本文在设计异常检测方法时，结合了每类攻击的特点，建立入侵检测系统模型。

2　PCA-GSA-AFSA-ELM 的工控网络异常行为预测模型

2.1　GSA-AFSA 算法

2.1.1　觅食行为

人工鱼 X_i 按式(1)在视野 *visual* 内随机产生一个状态 X_j[12]。

$$X_j = X_i + (2 \cdot rand - 1) \cdot visual \tag{1}$$

存在状态 X_1, X_2, …, X_v ($v \le try_number$) 的适应值

$f_k > f_i (k=1,2,\ldots,v)$。其中，$f_k(k=1,2,\ldots,v)$的最大适应值为$f(X_{best})$。下面对各个状态$X_1, X_2,\ldots,X_v(v\leq try_number)$进行惯性质量计算：

$$\begin{cases} m_k(X)=\dfrac{f_k(X)-f(X_i)}{f(X_{best})-f(X_i)} \\ M_k(X)=m_k(X)/\sum_{j=1}^{v} m_j(X) \end{cases} \quad (k=1,2,\cdots,v) \tag{2}$$

这样状态X_k的适应值$f(X_k)$的大小决定了惯性质量$M_k(X)$的大小，适应值越大，粒子惯性质量越大，粒子越趋向于最优解。以粒子惯性质量作为权重，计算加权后的中心位置。

$$X_c=\sum_{k=1}^{v} m_k(X)\cdot X_k \tag{3}$$

状态X_k的适应值$f(X_k)$越大，加权后的中心位置越偏向X_k。如果$f(X_c)>f(X_t)$，人工鱼X_t按照式(4)向下一个状态X_{next}移动。

$$X_{next}=X_i+\frac{X_c-X_i}{||X_c-X_i||}\cdot rand()\cdot step \tag{4}$$

相反，随机尝试次数try_number后，如若仍然不满足前进条件，则执行随机行为。

2.1.2 聚群行为

当前人工鱼X_t在视野$visual$范围内探索的伙伴数目为p，对其伙伴$X_j(j=1,2,\ldots,p)$按照式(5)计算权重，得到加权后的中心位置。

$$X_c=\sum_{j=1}^{p} m_j(X)\cdot X_j \tag{5}$$

如果$f(X_c)>f(X_t)$，人工鱼X_t按照式(6)向下一个状态X_{next}移动。

$$X_{next}=X_i+\frac{X_c-X_i}{||X_c-X_i||}\cdot rand()\cdot step \tag{6}$$

相反，随机尝试次数try_number后，如若仍然不满足前进条件，则执行随机行为。

2.2 GSA-AFSA 算法优化的 ELM

由于 ELM 初始输入权值和阈值是随机确定的，训练的精度和时间都会受随机性的影响，所以需要采用 GSA-AFSA 算法对 ELM 初始输入权值和阈值进行优化，从而避免盲目训练人工神经网络。

在本文中用 GSA-AFSA 算法优化极限学习机中的权值和阈值，建立工业控制网络通信异常行为检测的 GSA-AFSA-ELM 模型。该算法的具体步骤[13]如下：

a)算法参数初始化设置。鱼群规模 N、视野 visual、步长 step、最大迭代次数 T、尝试次数 try-number、拥挤度δ等。

b)初始化种群。将极限学习机的输入权值和阈值作为 GSA-AFSA 算法的人工鱼，长度为 D=k×(n+1)。其中，隐含层节点数目 k,输入向量维数 n。$\boldsymbol{\theta}^m$ 为种群中的第 m(1≤m≤popsize)个人工鱼。

$$\boldsymbol{\theta}^m=[w_{11}^m,w_{12}^m,\cdots,w_{1k}^m,w_{21}^m,w_{22}^m,\cdots,w_{2k}^m,\cdots,w_{n1}^m, w_{n2}^m,\cdots,w_{nk}^m,b_1^m,b_2^m,\cdots,b_k^m] \tag{7}$$

其中: w_{ij}^m、b_j^m 为[-Xmax,Xmax]中的随机数，一般 Xmax=1。将 ELM 的训练样本的均方根误差函数作为适应度函数，计算人工鱼的适应度值，并将最优个体的位置及食物浓度记录在公告板上。

c)人工鱼的行为选择。

(a)觅食行为。人工鱼 Xt 按式(1)在视野 visual 内随机产生一个状态 Xj。将比当前位置及食物浓度更高的粒子按式(2)计算权重，同时按照式(3)获得加权后的中心位置 Xc。如果中心位置的食物浓度相比于当前位置的食物浓度更高且不拥挤，则向前移动一步；相反，执行随机行为。

(b)聚群行为。当前人工鱼 Xt 在视野 visual 范围内探索的伙伴数目为 p，对其伙伴按照式(5)计算权重，得到加权后的中心位置 Xc。如果中心位置的食物浓度相比于当前位置的食物浓度更高且不拥挤，则向前移动一步；相反，执行觅食行为。

(c)追尾行为。当前人工鱼 Xt 在视野 visual 范围内探索食物浓度最高的伙伴并判断它的周围是否拥挤，如果成立，就向前一步；否则，执行觅食行为。

(d)随机行为。当前人工鱼 Xt 在视野 visual 范围内随机选择一个位置，向前一步。

d)更新公告板。将当前最优的人工鱼记录到公告板。

e)迭代次数加一，返回步骤 c)，直到迭代次数大于设置的最大迭代次数。

f)将公告板上的最优人工鱼输出代入式(8)中，利用 ELM 算法训练 SLFNs，输出网络通信指标预测值。

$$\boldsymbol{\beta}=\boldsymbol{H}^{+}\boldsymbol{T} \tag{8}$$

其中：$\boldsymbol{H}$ 为隐含层输出矩阵；$\boldsymbol{T}$ 为神经网络的输出；$\boldsymbol{\beta}$ 为输出权重；$\boldsymbol{H}^{+}$ 为广义逆。

GSA-AFSA 算法能有效地避免鱼群算法过早收敛到局部极值，提高了算法的收敛精度，算法的稳定性得到改善。

2.3 PCA-GSA-AFSA-ELM 工控网络异常行为预测模型

PCA-GSA-AFSA-ELM 预测模型工作流程如图 2 所示。

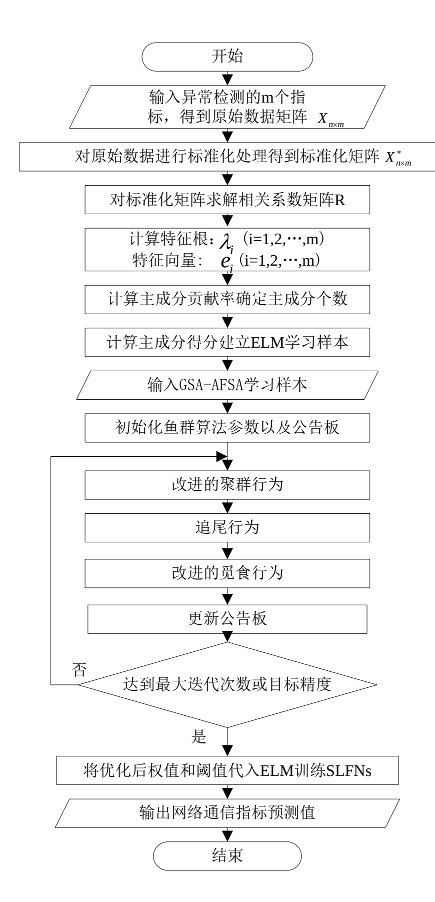


图 2　算法求解流程

3　测试与结果分析

3.1　算法有效性验证

3.1.1 实验数据描述

实验选用的数据集为异常检测领域广泛采用的 KDD99 数据集。训练集和测试集分别包含 494 021 条和 311 029 条数据。其中包括正常数据和攻击数据，主要包括 PROBE(probing attack)、DOS(denial of service attack)、U2R(user-to-root attack)和 R2L(remote-to-login attack) 四大类攻击。每条记录包含 41 维特征，其中最后一列为标签属性。

3.1.2 实验环境和参数设定

测试实验硬件为 Intel core i3-5005U CPU 2.00 GHz，操作系统为 Windows7 旗舰版，软件为 MATLAB 2015b，Python 2.7.13。

首先，上文根据工业控制系统通信信息进行入侵检测特征选取，利用 python 软件对 494 021 个样本 12 个特性[14]进行主成分分析。通过计算累计贡献率即可最终确定提取主成分个数。提取主成分之后，由公式计算得到主成分载荷及主成分得分。比较每一个主成分对应的各个原始特性的载荷，载荷越大，对应的主成分反映的该原始特性的信息量就越大。最后提取了 10 个主成分反映的主要原始特性信息。相应用 python 实现 PCA 的伪代码如下:

```
import numpy as np
import pandas as pd
from sklearn.decomposition import PCA

file=pd.read_csv('filename',sep=',',names=names,low_memory=False)
```

表 1 为 10 个主成分反映的主要原始特性。

表 1　主成分反映的主要原始特性

主成分	反映主要原始变量
1	连接持续时间
2	协议类型
3	目标主机的网络服务类型
4	从源主机到目标主机的数据字节数
5	从目标主机到源主机的数据字节数
6	是否成功登录
7	过去 2 s 内，与当前连接具有相同服务的连接数
8	过去 2 s 内，与当前连接具有相同的目标主机的连接数
9	前 100 个连接中，与当前连接具有相同目标主机相同源端口的连接所占的百分比
10	前 100 个连接中，与当前连接具有相同的目标主机相同服务的连接数

在 MATLAB 环境下，利用 GSA-AFSA-ELM 算法训练神经网络。选择应用最为广泛的 3 层单向前馈型神经网络，输入层神经元个数为 10，隐含层神经元个数为 30，输出层神经元为 5，分别为 Normal 正常数据、DoS 攻击、U2R 攻击、R2L 攻击、Probe 攻击。对应的神经网络拓扑如图 3 所示。

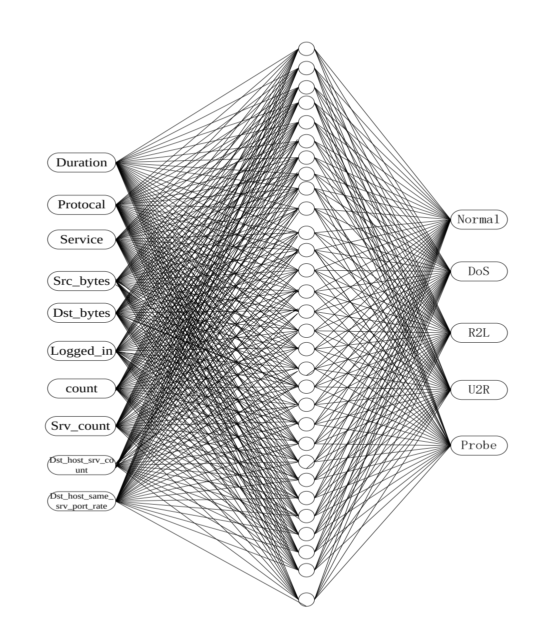


图 3　神经网络拓扑图

本文鱼群寻优部分迭代最大次数为 100 次，每次迭代过程中记录其鱼群适应度方差。适应度方差曲线比较如图 4 所示。

从图 4 中可以看出，AFSA-ELM 的方差在迭代 20 次陷入最小值，GSA-AFSA-ELM 迭代 81 次寻到最优解，跳出了局部最优陷阱。

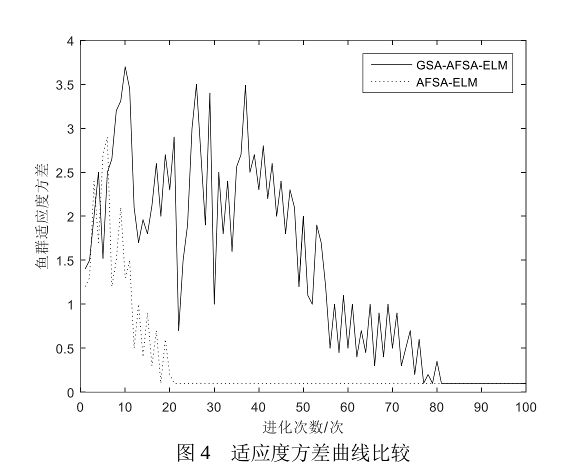


图 4　适应度方差曲线比较

图 5 所示为 AFSA-ELM 与 GSA-AFSA-ELM 算法性能对比。本文提出的 GSA-AFSA-ELM 算法训练 23 次时达到误差最小值，其值为 0.05；然而 AFSA-ELM 算法训练 36 次时达到误差最小值，其值为 0.15。因此在本文中提出的 GSA-AFSA-ELM 算法提高了收敛的精度，有效避免了早期收敛。

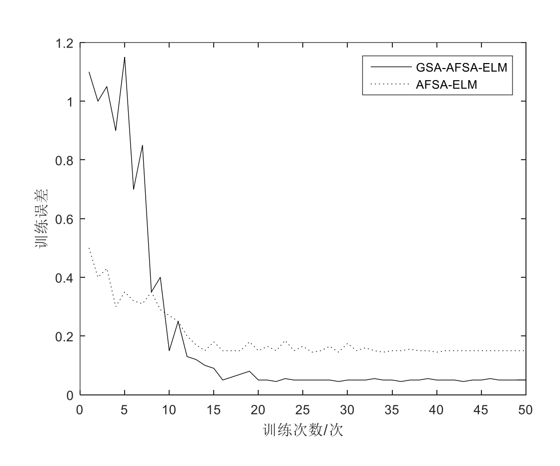
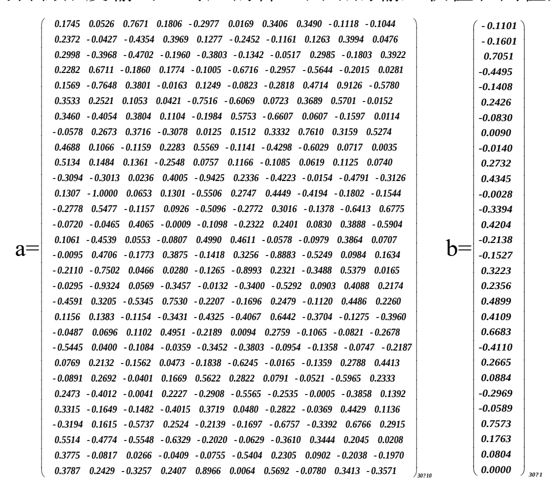


图 5　训练误差比较

经 GSA-AFSA 算法寻优后，将公告板上最优的人工鱼的位置及食物浓度输出，对应的神经网络的输入权值和阈值为



$$
a=\begin{pmatrix}
0.1745 & 0.0526 & 0.7671 & 0.1806 & -0.2977 & 0.0169 & 0.3406 & 0.3490 & -0.1118 & -0.1044 \\
0.2372 & -0.0427 & -0.4354 & 0.3969 & 0.1277 & -0.2452 & -0.1161 & 0.1263 & 0.3994 & 0.0476 \\
0.2998 & -0.3968 & -0.4702 & -0.1960 & -0.3803 & -0.1342 & -0.0517 & 0.2985 & -0.1803 & 0.3922 \\
0.2282 & 0.6711 & -0.1860 & 0.1774 & -0.1005 & -0.6716 & -0.2957 & -0.5644 & -0.2015 & 0.0281 \\
0.1569 & -0.7648 & 0.3801 & -0.0163 & 0.1249 & -0.0823 & -0.2818 & 0.4714 & 0.9126 & -0.5780 \\
0.3533 & 0.2521 & 0.1053 & 0.0421 & -0.7516 & -0.6069 & 0.0723 & 0.3689 & 0.5701 & -0.0152 \\
0.3460 & -0.4054 & 0.3804 & 0.1104 & -0.1984 & 0.5753 & -0.6607 & 0.0607 & -0.1597 & 0.0114 \\
-0.0578 & 0.2673 & 0.3716 & -0.3078 & 0.0125 & 0.1512 & 0.3332 & 0.7610 & 0.3159 & 0.5274 \\
0.4688 & 0.1066 & -0.1159 & 0.2283 & 0.5569 & -0.1141 & -0.4298 & -0.6029 & 0.0717 & 0.0035 \\
0.5134 & 0.1484 & 0.1361 & -0.2548 & 0.0757 & 0.1166 & -0.1085 & 0.0619 & 0.1125 & 0.0740 \\
-0.3094 & -0.3013 & 0.0236 & 0.4005 & -0.9425 & 0.2336 & -0.4223 & -0.0154 & -0.4791 & -0.3126 \\
0.1307 & -1.0000 & 0.0653 & 0.1301 & -0.5506 & 0.2747 & 0.4449 & -0.4194 & -0.1802 & -0.1544 \\
-0.2778 & 0.5477 & -0.1157 & 0.0926 & -0.5096 & -0.2772 & 0.3016 & -0.1378 & -0.6413 & 0.6775 \\
-0.0720 & -0.0465 & 0.4065 & -0.0009 & -0.1098 & -0.2322 & 0.2401 & 0.0830 & 0.3888 & -0.5904 \\
0.1061 & -0.4539 & 0.0553 & -0.0807 & 0.4990 & 0.4611 & -0.0578 & -0.0979 & 0.3864 & 0.0707 \\
-0.0095 & 0.4706 & -0.1773 & 0.3875 & -0.1418 & 0.3256 & -0.8883 & -0.5249 & 0.0984 & 0.1634 \\
-0.2110 & -0.7502 & 0.0466 & 0.0280 & -0.1265 & -0.8993 & 0.2321 & -0.3488 & 0.5379 & 0.0165 \\
-0.0295 & -0.9324 & 0.0569 & -0.3457 & -0.0132 & -0.3400 & -0.5292 & 0.0903 & 0.4088 & 0.2174 \\
-0.4591 & 0.3205 & -0.5345 & 0.7530 & -0.2207 & -0.1696 & 0.2479 & -0.1120 & 0.4486 & 0.2260 \\
0.1156 & 0.1383 & -0.1154 & -0.3431 & -0.4325 & -0.4067 & 0.6442 & -0.3704 & -0.1275 & -0.3960 \\
-0.0487 & 0.0696 & 0.1102 & 0.4951 & -0.2189 & 0.0094 & 0.2759 & -0.1065 & -0.0821 & -0.2678 \\
-0.5445 & 0.0400 & -0.1084 & -0.0359 & -0.3452 & -0.3803 & -0.0954 & -0.1358 & -0.0747 & -0.2187 \\
0.0769 & 0.2132 & -0.1562 & 0.0473 & -0.1838 & -0.6245 & -0.0165 & -0.1359 & 0.2788 & 0.4413 \\
-0.0891 & 0.2692 & -0.0401 & 0.1669 & 0.5622 & 0.2822 & 0.0791 & -0.0521 & -0.5965 & 0.2333 \\
0.2473 & -0.4012 & -0.0041 & 0.2227 & -0.2908 & -0.5565 & -0.2535 & -0.0005 & -0.3858 & 0.1392 \\
0.3315 & -0.1649 & -0.1482 & -0.4015 & 0.3719 & 0.0480 & -0.2822 & -0.0369 & 0.4429 & 0.1136 \\
-0.3194 & 0.1615 & -0.5737 & 0.2524 & -0.2139 & -0.1697 & -0.6757 & -0.3392 & 0.6766 & 0.2915 \\
0.5514 & -0.4774 & -0.5548 & -0.6329 & -0.2020 & -0.0629 & -0.3610 & 0.3444 & 0.2045 & 0.0208 \\
0.3775 & -0.0817 & 0.0266 & -0.0409 & -0.0755 & -0.5404 & 0.2305 & 0.0902 & -0.2038 & -0.1970 \\
0.3787 & 0.2429 & -0.3257 & 0.2407 & 0.8966 & 0.0064 & 0.5692 & -0.0780 & 0.3413 & -0.3571
\end{pmatrix}_{30?10}
\quad
b=\begin{pmatrix}
-0.1101 \\ -0.1601 \\ 0.7051 \\ -0.4495 \\ -0.1408 \\ 0.2426 \\ -0.0830 \\ 0.0090 \\ -0.0140 \\ 0.2732 \\ 0.4345 \\ -0.0028 \\ -0.3394 \\ 0.4204 \\ -0.2138 \\ -0.1527 \\ 0.3223 \\ 0.2356 \\ 0.4899 \\ 0.4109 \\ 0.6683 \\ -0.4110 \\ 0.2665 \\ 0.0884 \\ -0.2969 \\ -0.0589 \\ 0.7573 \\ 0.1763 \\ 0.0804 \\ 0.0000
\end{pmatrix}_{30?1}
$$

3.1.3 实验结果分析

算法训练的次数对算法的检测准确率存在影响，文中对比了 GSA-AFSA-ELM、AFSA-ELM 和 ELM 这三种算法，如表 2 所示，本文方法在训练次数和精确程度上都有明显优势。

表 2　各方法的训练次数及对应的检测率

方法	训练次数	检测率/%
本文方法	22	80.07
AFSA-ELM	36	70.13
ELM	47	68.27

现将该模型与 AFSA-ELM、ELM、BP 模型进行对比研究，利用训练集和测试集对其进行训练，并产生预测结果。对比结果如图 6 所示。

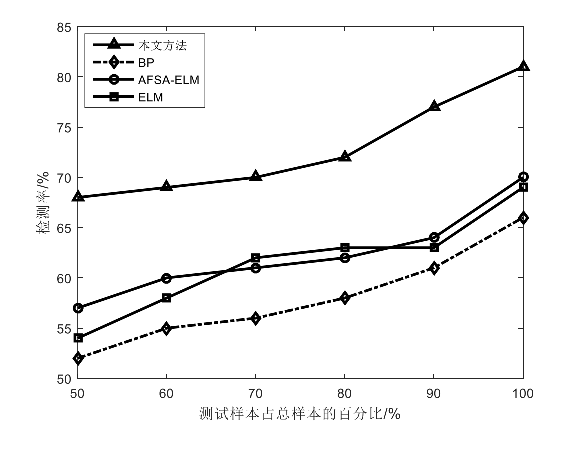


图 6　四种算法检测率比较

相比于 BP 神经网络，ELM 极限学习机算法学习速度快、泛化性能好。图 7 为四种模型训练时间的对比。经过 GSA-AFSA 算法优化后的 ELM 训练时间为 3.421 9 s，而单纯的 ELM 训练的时间为 3.468 8 s，产生训练时间上的差距主要由于单纯的 ELM 训练过程输入权值和阈值是随机的，为提升预测结果的精度和决定系数，需要反复训练寻找满足要求的参数。这种较为盲目随的训练方法增长了训练时间。

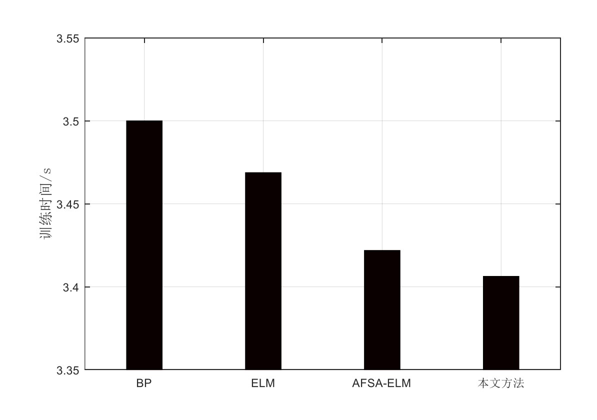


图 7　多种模型训练时间对比

四种模型对应的测试时间如图 8 所示。

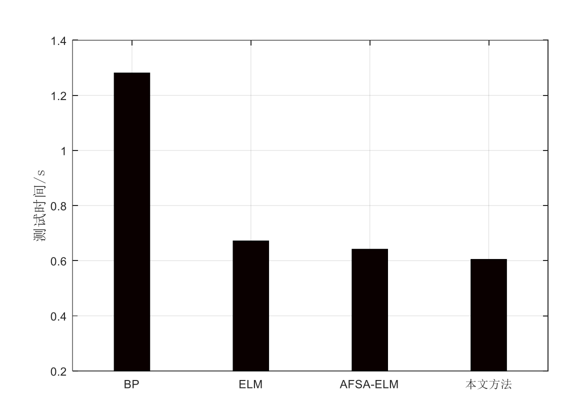


图 8　多种模型测试时间对比

图 9 为本文实验 ELM、AFSA-ELM、GSA-AFSA-ELM 这三种系统分别对 DoS、U2R、R2L、Probe 攻击的检测率，其中 DoS 攻击的检测率最高，可以达到 96.90%以上，对于 R2L 的检测率达到 86.27%。

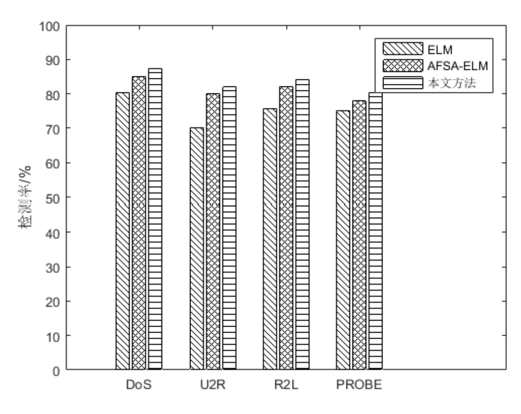


图 9　三种系统检测率对比

3.2　算法性能分析

3.2.1 实验数据描述

为了验证本文算法在工业控制领域的适用性，本文以从某风力发电场获取的数据为基础，对本文算法进行性能分析。图 10 为某风力发电场的数据采集与同步系统架构拓扑图。

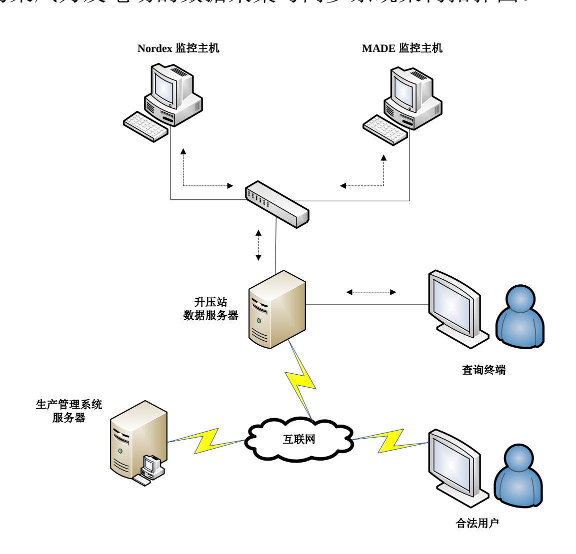


图 10　数据采集与同步系统架构拓扑图

表 3 为某风力发电场 web 发布的部分数据。

表 3　部分采集数据

风机编号	风机名称	额定功率(KW)	当前风速(m/s)	当前功率(KW)
1	MADE-01号风机	660KW	7.9	134.6
2	MADE-02号风机	660KW	5.9	0.0
3	MADE-03号风机	660KW	9.9	151.3
4	MADE-04号风机	660KW	7.6	123.2
5	MADE-05号风机	660KW	6.9	65.7
6	MADE-06号风机	660KW	6.4	0.0
7	MADE-07号风机	660KW	6.7	99.0
8	MADE-08号风机	660KW	0.0	0.0
9	MADE-09号风机	660KW	7.6	59.7
10	NORDEX-10号风机	600KW	7.4	276.33
11	NORDEX-11号风机	600KW	7.0	93.55
12	NORDEX-12号风机	600KW	6.7	80.07
13	NORDEX-13号风机	600KW	5.3	47.31
14	NORDEX-14号风机	600KW	7.3	156.7

3.2.2 实验参数设定及数据分析

从某风力发电场获取数据 147 条，其中正常数据为 98 条，异常数据为 49 条。首先利用 python 软件对 147 个样本 30 个特性进行主成分分析。通过计算最终确定提取 10 个特性作为主成分。构造单隐层前馈神经网络，将 10 个主成分作为输入层节点的输入，输出层为 5 个节点，分别为正常数据、DoS 攻击、中间人攻击、内部攻击和端口扫描。其余参数设置与上文相同。

对于在工控系统采集到的 147 条数据，其中 102 条作为训练集，训练集包含 54 条正常数据和 48 条异常数据；45 条作为测试集，测试集中包含 31 条正常通信数据和 17 条异常通信数据。

现将本文方法与 AFSA-ELM、ELM、BP 网络进行对比，利用工控系统采集到的数据对其进行训练，并产生预测结果，对比结果见图 11。

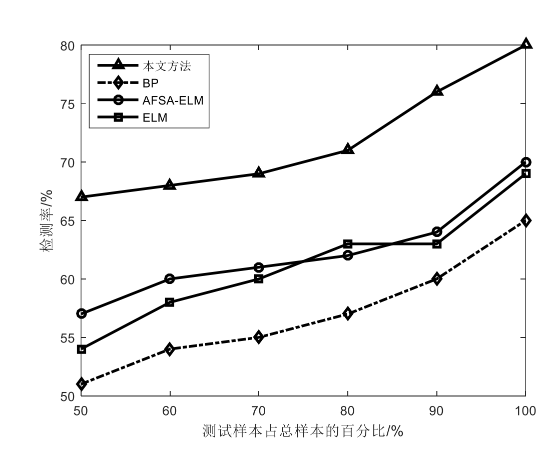


图 11　四种算法检测率比较

本文实验 ELM、AFSA-ELM、本文方法这三种系统分别对 DoS、中间人、内部、端口扫描攻击的检测率，其中 DoS 攻击的检测率最高，可以达到 87.31%以上，对于内部攻击的检测率达到 84.24%。图 12 为这三种系统对各个攻击的检测率。

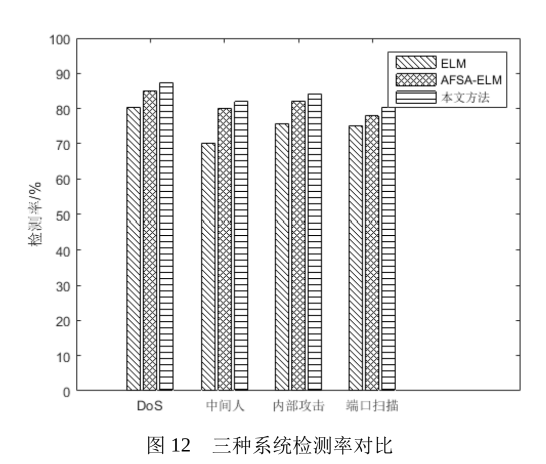


图 12 三种系统检测率对比

4 结束语

本文针对工业控制网络入侵检测问题提出一种主成分分析及 GSA-AFSA-ELM 神经网络的通信异常检测方法。根据本系统的工作流程可知该方法有以下优势：主成分分析用于提取数据的主要特征分量，常用于高维数据的降维，简化复杂问题。ELM 极限学习机相比于 BP 神经网络，学习速度快、泛化性能好。在训练过程中 ELM 算法无须调整，只需设置隐含层神经元的个数，便可获得唯一的最优解。利用 GSA-AFSA-ELM 方法可实现异常通信行为的预测。实验结果表明，本文算法能够有效地检测网络数据中的异常行为，异常检测的准确性提升了 3.51%，有效地缩短了 0.609 4 s 的检测时间，实现了利用深度学习预测通信行为的能力。下一步将研究采用模糊逻辑实现入侵检测机制，并重点提升检测系统的容错性。

参考文献：

[1] 尚文利，安攀峰，万明，等. 工业控制系统入侵检测技术的研究及发展综述 [J]. 计算机应用研究, 2017, 34 (2): 328-333+342. (Shang Wen Li, An Pan Feng, Wan Ming, *et al.* Research and development overview of intrusion detection technology in industrial control system [J]. Journal of Application Research of Computers, 2017, 34 (2): 328-333+342.)

[2] 赖英旭，刘增辉，蔡晓田，等. 工业控制系统入侵检测研究综述 [J]. 通信学报, 2017, 38 (2): 143-156. (Lai YingXu, Liu Zenhui, Cai X iaotian, *et al.* Research on intrusion detection of industrial control system [J]. Journal of Communication, 2017, 38 (2): 143-156.)

[3] 尚文利，李琳，万明，等. 基于优化单类支持向量机的工业控制系统入侵检测算法 [J]. 信息与控制, 2015, 44 (6): 678-684. (Shang Wenli, Li Lin, Wan Ming, *et al.* Intrusion detection algorithm based on optimized one-class support vector machine for industrial control system [J]. Information and Control, 2015, 44 (6): 678-684.)

[4] Yang Yi, McLaughlin K, Littler T, *et al.* Intrusion detection system for IEC 60870-5-104 based SCADA networks [C]// Proc of Power and Energy Society General Meeting.

[5] Piscataway, NJ: IEEE Press, 2013: 1-5.

[6] 侯重远，江汉红，芮万智，等. 工业网络流量异常检测的概率主成分分析法 [J]. 西安交通大学学报, 2012, 46 (2): 70-75. (Hou Chongyuan, Jiang Hanhong, Rui Wanzhi, *et al.* A probabilistic principal component analysis approach for detecting traffic anomaly in industrial networks [J]. Journal of XI’AN JiaoTong University, 2012, 46 (2): 70-75.)

[7] Zhou C, Huang S, Xiong N, *et al.* Design and analysis of multimodel-based anomaly intrusion detection system in industrial process automation [J]. IEEE Trans on System, Man, and Cybernetics, 2015, 45 (10): 2168-2216.

[8] Aburomman A A, Reaz M B I. A novel SVM-KNN-PSO ensemble method for intrusion detection system [J]. Applied Soft Computing, 2016, 38 (1): 360-372.

[9] Parvania M, Koutsandria G, Muthukumar V, *et al.* Hybrid control network intrusion detection systems for automated power distribution systems [C]// Proc of IEEE//IFIP International Conference on Dependable Systems and Networks. [S. l.] : IEEE Computer Society, 2014: 774-779.

[10] Jensen P T. Detection and characterization of port scan attacks [J]. Key Engineering Materials, 2014, 602-603 (3): 93-96.

[11] Ponomarev S, Atkison T. Industrial control system network intrision detection by telemetry analysis [J]. IEEE Trans on Dependable and Secure Computing, 2016, 13 (2): 252-260.

[12] Bishop M, Gates C. Defining the insider threat [C]// Proc of Workshop on Cyber Security & Information Intelligence Research: Developing Strategies To Meet the Cyber Security & Information Intelligence Challenges Ahead. 2008: 1-3.

[13] 冯国明，郭承军，叶晶晶. 基于万有引力搜索算法改进的人工鱼群算法 [J]. 数学学习与研究, 2016 (1): 144-146. (Feng Guoming, Guo Chengjun, Ye Jingjing. Artificial fish swarm algorithm improved based on universal gravitational search algorithm [J]. Mathematics Learning and Research, 2016 (1): 144-146.)

[14] 周华平，袁月. 改进鱼群算法优化的 ELM 在乳腺肿瘤辅助诊断中的应用研究 [J]. 计算机工程与科学, 2017, 39 (11): 2144-2152. (Zhou Huaping, Yuan Yue. Application of ELM in computer-aided diagnosis of breast tumors based on improved fish swarm optimization algorithm [J]. Computer Engineering & Science, 2017, 39 (11): 2144-2152.)

[15] 赵新星，姜青山，陈路莹，等. 一种面向网络入侵检测的特征选择方式 [A]// 第 26 届中国数据库学术会议论文集 (B 辑) [C]// 中国计算机学会数据库专业委员会, 2009: 488-493. (Zhao Xinxing, Jiang Q ingshan, Chen Luying, *et al.* A feature selection method for network intrusion detection [A]// Proc of the 26th China Database Conference Papers (series B) [C]// China Computer Society Database Specialized Committee. 2009: 488-493.)